



## ORIGINAL RESEARCH

# Inter-rater Reliability Assessment of ASPECT-R (A Study Pragmatic-Explanatory Characterization Tool-Rating)

**FUNDING:** Financial support for the development of the ASPECT-R instrument and this study was provided by Janssen Scientific Affairs, LLC.

**FINANCIAL DISCLOSURES:** Drs. Bossie, Alphs, Williamson, and Mr. Kurut are employees of Janssen Scientific Affairs, LLC, Titusville, New Jersey, USA. Dr. Mao is an employee of Janssen Research & Development, Titusville, New Jersey, USA.

**ADDRESS CORRESPONDENCE TO:**

Larry Alphs, MD, PhD, Medical Affairs, Central Nervous System, Janssen Scientific Affairs, LLC, 1125 Trenton-Harbourton Road, Titusville, NJ 08560; Phone: 609-703-2177; Fax: 609-730-3232; E-mail: lalphs@its.jnj.com

**KEY WORDS:** ASPECT-R, Trial Design, Pragmatic, Explanatory, Inter-rater reliability, Intraclass correlation coefficient, real-world

by **CYNTHIA A. BOSSIE, PhD; LARRY D. ALPHS, MD, PhD; DAVID WILLIAMSON, PhD; LIAN MAO, PhD; CLENNON KURUT, BS; and THE ASPECT-R RATER TEAM**

*Drs. Bossie, Alphs, and Williamson are with Janssen Scientific Affairs, LLC, Titusville, New Jersey, USA; Dr. Mao and Mr. Kurut are with Janssen Research & Development, Titusville, New Jersey, USA*

*Innov Clin Neurosci.* 2016;13(3-4):27-31

## ABSTRACT

**Objective:** The increasing importance of real-world data for clinical and policy decision making is driving a need for close attention to the pragmatic versus explanatory features of trial designs. ASPECT-R (A Study Pragmatic-Explanatory Characterization Tool-Rating) is an instrument informed by the PRECIS tool, which was developed to assist researchers in designing trials that are more pragmatic or explanatory. ASPECT-R refined the PRECIS domains and includes a detailed anchored rating system. This analysis established the inter-rater reliability of ASPECT-R.

**Design:** Nine raters (identified from a convenience sample of persons knowledgeable about psychiatry clinical research/study design) received ASPECT-R training materials and 12 study publications. Selected studies assessed antipsychotic treatment in schizophrenia, were published in peer-reviewed journals, and represented a range of studies across a pragmatic-explanatory continuum as determined by authors (CB/LA). After completing training, raters reviewed the 12 studies and rated the study domains

using ASPECT-R. Intraclass correlation coefficients were estimated for total and domain scores. Qualitative ratings then were assigned to describe the inter-rater reliability.

**Results:** ASPECT-R scores for the 12 studies were completed by seven raters. The ASPECT-R total score intraclass correlation coefficient was 0.87, corresponding to an excellent inter-rater reliability. Domain intraclass correlation coefficients ranged from 0.85 to 0.31, corresponding to excellent to poor inter-rater reliability.

**Conclusion:** The inter-rater reliability of the ASPECT-R total score was excellent, with excellent to good inter-rater reliability for most domains. The fair to poor inter-rater reliability for two domains may reflect a need for improved domain definition, anchoring, or training materials. ASPECT-R can be used to help understand the pragmatic-explanatory nature of completed or planned trials.

## INTRODUCTION

Clinical trials have characteristics that are pragmatic (i.e., effectiveness studies asking whether an

intervention works under usual or real-world conditions) and/or explanatory (i.e., asks whether an intervention works under highly controlled and well-defined conditions). Trials utilizing either or both of these types of characteristics have value, with neither being intrinsically superior to the other. However, the increasing importance of real world data for clinical and policy decision making is driving a need for close attention to the pragmatic-explanatory characteristics or continuum of the study design.

The ASPECT-R (A Study Pragmatic-Explanatory Characterization Tool-Rating; ©2014 Janssen Pharmaceuticals, Inc.) is an instrument informed from the PRECIS (Pragmatic-Explanatory Continuum Indicator Summary) tool described previously by Thorpe et al.<sup>1</sup> The PRECIS, with its 10 domains assessing study design on an unmarked visual analog scale, was developed to assist researchers when designing trials along the explanatory:pragmatic spectrum.<sup>1</sup> Tosh et al.<sup>2</sup> adapted the PRECIS tool by adding a 6-point visual analog scale to each domain and termed this instrument the Pragmascope. An inter-rater reliability (IRR) assessment of the Pragmascope has been performed<sup>3</sup> and a modified version of the PRECIS instrument (the PRECIS-2) is under Phase III evaluation for its validity and reliability.<sup>3</sup>

As opposed to the 10-domains of the PRECIS and Pragmascope instruments, the ASPECT-R tool assesses six study design domains, with these domains specifically related to the explanatory:pragmatic spectrum. The four domains excluded when developing the ASPECT-R tool were those considered to be redundant or more focused on measures of study quality. Each of the six ASPECT-R domains are rated on a 0 to 6 scale where 0=extremely explanatory and 6=extremely pragmatic, with a 0 to 36 range for the ASPECT-R total

score. The ASPECT-R also includes detailed definitions of terms and descriptive anchors to facilitate greater reliability across raters.<sup>4</sup> The developers of the ASPECT-R tool have recently described the tool and its use in evaluating published clinical trials.<sup>5</sup> The ASPECT-R tool is considered useful in the study design stage as well as to assess the explanatory versus pragmatic nature of published trials.

The primary objective of this study was to evaluate the IRR of the ASPECT-R total score, with IRR assessments of each domain score being secondary objectives.

## MATERIALS AND METHODS

**Studies.** Twelve studies for the IRR assessment of the ASPECT-R were identified through a literature search of comparative studies of interventions applied in a clinical setting. Full citations of the studies utilized are listed in Table 1. The studies assessed antipsychotic treatment in schizophrenia (e.g., long-acting injectable antipsychotics and oral antipsychotics), were published in peer-reviewed journals between January 1993 and July 2013, enrolled at least 100 subjects, and represented a range of studies across a pragmatic-explanatory design continuum as assessed by the authors (CB and LA).

Each of the 12 studies was reviewed by ASPECT-R developers (CB and LA) with a consensus rating between the developers determined for each domain of each study.

**Raters.** Nine ASPECT-R raters were identified from a convenience sample of individuals with knowledge of clinical research and study designs in the area of psychiatry and the treatment of schizophrenia. Raters were not required to have advanced degrees; however, they were to have a level of expertise regarding the study's population of interest (i.e., schizophrenia) as it related to the epidemiology, clinical characteristics, course of illness,

treatment regimens and modalities, and clinical response.

All nine raters were provided with a concise ASPECT-R training package consisting of a slide deck and the Excel® worksheet files. They were given a 30-minute orientation regarding the project and directions on the scope of their participation. The raters were allowed 12 weeks to complete their review of each of the 12 studies and rate them using ASPECT-R. The completed ASPECT-R ratings were returned to the team statistician (author LM) for analysis.

**Statistical methods.** Sample size estimation for this IRR study consists of determining the number of raters and studies needed to achieve an acceptable level of precision for estimating intraclass correlation coefficients (ICCs). It was anticipated that the ICCs for the ASPECT-R total score would be 0.75 or greater; therefore, 10 or more studies and eight raters would provide an ICC estimate with a 95-percent confidence interval (CI) within  $\pm 0.20$  of the observed ICCs.

ICCs for the total score of the ASPECT-R tool, as well as for each domain score were estimated. Cicchetti et al.<sup>6</sup> provide commonly cited cutoffs for qualitative ratings of agreement based on ICC values with IRR being excellent for values between 1.0 and 0.75, good for values between 0.74 and 0.60, fair for values between 0.59 and 0.40, and poor for values less than 0.40.

Regression and correlation analyses were performed for the ASPECT-R total scores assigned by the raters versus those of the developers.

## RESULTS

ASPECT-R scores were available for all 12 studies by seven of the nine raters. The ASPECT-R total score ICC was 0.87, corresponding to an excellent IRR rating (i.e., a score above 0.75). As illustrated in Figure 1, ICC scores of four of the six individual domains (Intervention Flexibility:

Experimental/Comparison; Medical Practice Setting/Practitioner Expertise: Experimental/Comparison, Follow-up Intensity/Duration, and Participant Compliance) ranged from 0.85 (IRR=excellent) to 0.64 (IRR=good), with the domain of Participant Eligibility having an ICC of 0.55 (IRR=fair) and Primary Trial Outcome(s) having an ICC of 0.31 (IRR=poor). A strong correlation was found between the rater and the developer ASPECT-R total scores with a Pearson correlation of 0.92 ( $R^2=0.85$ ).

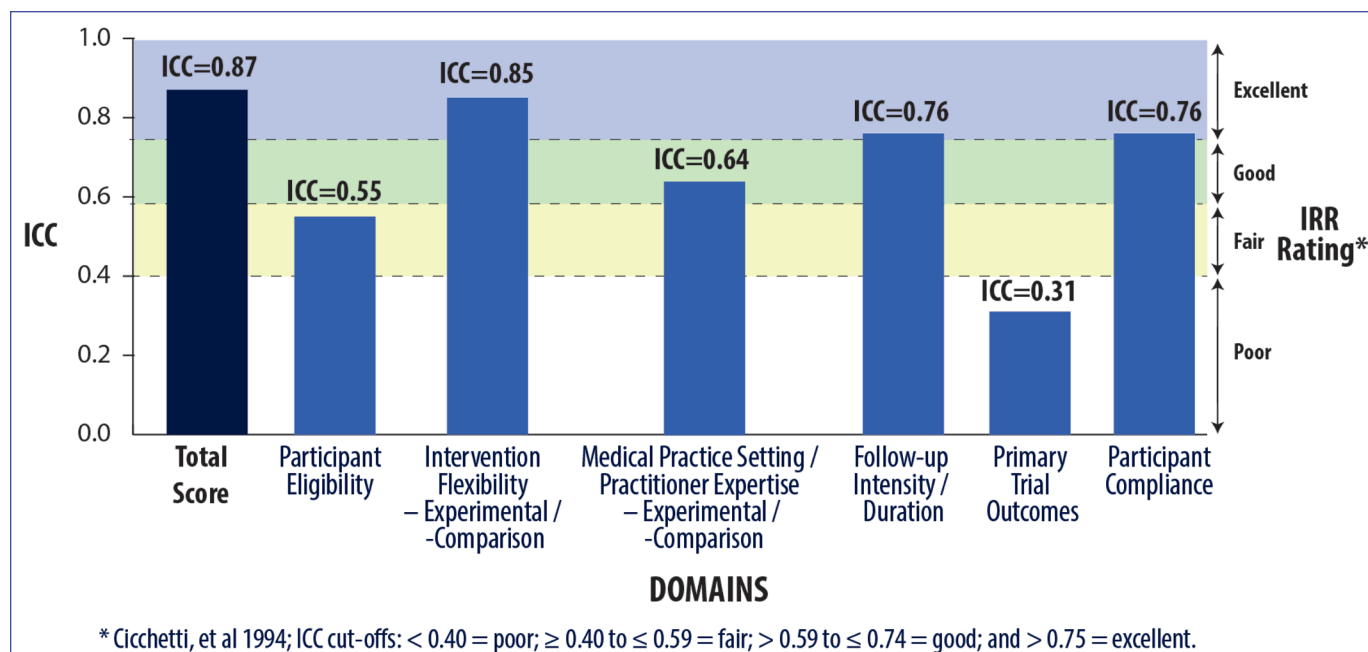
The ASPECT-R total score mean, minimum, maximum, median, and standard deviation values assigned by the team of raters for each study and those of the developer consensus are summarized in Table 2.

## DISCUSSION

This analysis establishes that the IRR of the total score of the ASPECT-R (beta-version) tool is excellent. Assessments of the six individual domain scores suggest that most had IRRs that were good to excellent. The lower IRRs noted for the domain of participant

**TABLE 1.** ASPECT-R inter-rater reliability assessment study citations

- Bitter I, et al. Comparative effectiveness of depot and oral second generation antipsychotic drugs in schizophrenia: a nationwide study in Hungary. *Euro Neuropsychopharmacology* 2013;23:1883–1890.
- Gaebel W, et al. Relapse prevention in schizophrenia and schizoaffective disorder with risperidone long-acting injectable vs quetiapine: results of a long-term, open-label, randomized clinical trial. *Neuropsychopharmacology*. 2010;35:2367–2377.
- Grimaldi-Bensouda L, et al. Does long-acting injectable risperidone make a difference to the real-life treatment of schizophrenia? results of the Cohort for the General study of Schizophrenia (CGS). *Schizophr Res*. 2012;134(23):187–194.
- Kane JM, et al. Olanzapine long-acting injection: a 24-week, randomized, double-blind trial of maintenance treatment in patients with schizophrenia. *Am J Psychiatry*. 2010;167:181–189.
- Keks NA, et al. Long-acting injectable risperidone v. olanzapine tablets for schizophrenia or schizoaffective disorder. *Br J Psychiatry*. 2007;191:131–139.
- Macfadden W, et al. A prospective study comparing the long-term effectiveness of injectable risperidone long-acting therapy and oral aripiprazole in patients with schizophrenia. *Psychiatry (Edmont)*. 2010;7(11):23–31.
- Olivares JM, et al. Long-term outcomes in patients with schizophrenia treated with risperidone long-acting injection or oral antipsychotics in Spain: results from the electronic Schizophrenia Treatment Adherence Registry (e-STAR). *Euro Psych*. 2009;24:287–296.
- Pandina G, et al. A double-blind study of paliperidone palmitate and risperidone long-acting injectable in adults with schizophrenia. *Prog Neuro-psychopharmacol Biol Psychiatry*. 2011;15(35):218–226.
- Pandina G, et al. A randomized, placebo-controlled study to assess the efficacy and safety of 3 doses of paliperidone palmitate in adults with acutely exacerbated schizophrenia. *J Clin Psychopharmacol*. 2010;30(3):235–244.
- Rosenheck RA, et al. Long-acting risperidone and oral antipsychotics in unstable schizophrenia. *N Engl J Med*. 2011;364(9):842–851.
- Tiihonen J, et al. A nationwide cohort study of oral and depot antipsychotics after first hospitalization for schizophrenia. *Am J Psychiatry*. 2011;68:603–609.
- Zhu B, et al. Time to discontinuation of depot and oral first-generation antipsychotics in the usual care of schizophrenia. *Psychiatr Serv*. 2008;59:315–317.



**FIGURE 1.** Intraclass correlation coefficient scores and inter-rater reliability ratings of ASPECT-R total score and individual domains

**TABLE 2.** ASPECT-R total scores: \* developer consensus and raters' values (Pearson correlation=0.92; R<sup>2</sup>=0.85)

STUDY	DEVELOPER CONSENSUS	RATERS' VALUES				
		MEAN	MINIMUM	MAXIMUM	MEDIAN	STANDARD DEVIATION
Bitter, 2013	35	33.5	29.5	35	34	1.89
Gaebel, 2010	18	15.5	11	22	14	3.82
Grimaldi-Bensouda, 2012	29	30.2	21.5	35	31	4.47
Kane, 2010	9	10.9	4	18	11	4.56
Keks, 2007	13	12.1	6	19.5	11	4.78
Macfadden, 2010	12.5	15.9	12	23	14	4.63
Olivares, 2009	33	30.3	27.5	32	31	1.9
Pandina, 2010	5	4.7	2	10	3.5	3.2
Pandina, 2011	6	8.7	1	16	8	5.1
Rosenheck, 2011	13	16.7	12	22	16	3.8
Tiihonen, 2011	35	32.4	26.5	35	34	3.1
Zhu, 2008	35	29.6	27	32	30	1.7

\*Each of the six ASPECT-R domains was scored on a 0–6 scale; the range of ASPECT-R total score was 0 to 36.

eligibility (IRR of fair) and primary trial outcomes (IRR of poor) may reflect a need for improvements in domain definitions, modifications of the anchoring criteria, or an increased emphasis on these domains in the training materials. Thus, modifications were made to these two domains in the published tool.<sup>5</sup> That being said, although these results provide strong support for the utility of our overall score, it may be premature to draw firm conclusions about the specific domain scores that are more or less reliable. Subtest or domain scores are less stable than overall scores psychometrically, and our single study is not able to parse out the relative impact of specific domain definitions from the effects specific to our small group of raters or effects specific to this sample of

publications. Replication with larger samples of raters with different collections of publications will provide greater confidence in which domains are scored most consistently.

The establishment of the IRR for the ASPECT-R tool, or any other tool developed to assess the pragmatic versus explanatory nature of a clinical trial, may be limited by factors such as poorly documented or unavailable study-related design or study conduct information. Optimally, the use of the ASPECT-R tool would be by raters that have considerable clinical trial expertise regarding the population of interest and its treatment. In this analysis, we utilized raters that had a level of expertise and knowledge of the schizophrenia population of interest as well as the use of antipsychotics

and psychotherapy in its treatment. Their knowledge of these factors as well as their experience in clinical trial design may have impacted their ratings and these findings.

## CONCLUSION

In conclusion, these findings indicate that the total score of the ASPECT-R has an excellent IRR. The use of the ASPECT-R has at least two clear applications. First, in the clinical trial design process it provides a clear descriptive framework for researchers to consistently identify where a planned study's key design domains lie along the pragmatic-explanatory continuum. Secondly, it allows healthcare providers and other researchers a way to interpret the findings of completed clinical trials and put them in a pragmatic to

explanatory continuum context when presented with inconsistent or conflicting trial results. It then offers a better understanding of the generalizability of the clinical trial results to real-world circumstances.

## ACKNOWLEDGMENTS

The authors acknowledge the contributions made by the ASPECT-R Rater Team: Hearee Chung, PharmD, Principal Scientific Affairs Liaison, Janssen Scientific Affairs, LLC; Peter Dorson, PharmD, Principal Scientific Affairs Liaison, Janssen Scientific Affairs, LLC; Dilesh Doshi, PharmD, Director, HECOR Translational Science, Janssen Scientific Affairs, LLC; Jennifer Kern Sliwa, PharmD, Director, Medical Information, Central Nervous System, Janssen Scientific Affairs, LLC; Branslav Mancevski, MD, Director, Clinical Development, Janssen Scientific Affairs, LLC; and Dean Najarian,

PharmD, Principal Scientific Affairs Liaison, Janssen Scientific Affairs, LLC.

Editorial support was provided by Susan Ruffalo, PharmD, MedWrite, Inc., Newport Coast, California.

These materials were presented at the 15th International Congress on Schizophrenia Research. March 28 to April 1, 2015.

## REFERENCES

1. Thorpe KE, Zwarenstein M, Oxman AD, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *J Clin Epidemiol*. 2009;62:464–475.
2. Tosh G, Soares-Weiser K, Adams CE. Pragmatic vs explanatory trials: the Pragmascope tool to help measure differences in protocols of mental health randomized controlled trials. *Dialogues Clin Neurosci*. 2011;13:209–215.
3. Loudon K, Zwarenstein M, Sullivan F, et al. Making clinical trials more relevant: improving and validating the PRECIS tool for matching trial design decisions to trial purpose. *Trials*. 2013;14:115.
4. Bossie CA, Alphas L. ASPECT-R: a tool to rate the pragmatic (effectiveness)/explanatory (efficacy) characteristics of a clinical trial design. Poster. 14th Intl Congress on Schizophrenia Research (ICOSR) April 21 to 25, 2013. Orlando, FL.
5. Alphas LA, Bossie CA. ASPECT-R: a tool to rate the pragmatic and explanatory characteristics of a clinical trial design. *Innov Clin Neurosci*. 2016;13(1–2):15–26.
6. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*. 1994;6(4):284–290. ■